EXHIBIT 6

[Submitting Counsel below]

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

IN RE: UBER TECHNOLOGIES, INC. PASSENGER SEXUAL ASSAULT LITIGATION	**MDL No. 3084 CRB**
This Document Relates to: ALL ACTIONS	**DECLARATION OF DOUGLAS FORREST IN SUPPORT OF PLAINTIFFS' BRIEF FOR PLAINTIFFS' PROPOSED ORDER GOVERNING THE PRODUCTION OF ELECTRONICALLY STORED INFORMATION AND HARD COPY DOCUMENTS**

DECLARATION OF DOUGLAS FORREST

I, Douglas Forrest, state and declare as follows:

1. I am the Senior Vice President, eDiscovery Analytics & Strategy, at International Litigation Services ("ILS"), which is located in Irvine, California (www.ilsteam.com). I have been retained as a consultant for the Plaintiffs in this action for my knowledge relating to Electronically Stored Information as it relates to creating an Electronically Stored Information Order. The facts stated in this declaration, except as otherwise explicitly noted, are within my own personal knowledge and, if called as a witness to testify, I could and would competently testify to the facts contained in this declaration.

2. I make this declaration in support of Plaintiffs' Proposed ESI Order. I am familiar with the issues raised with the competing ESI Orders, having closely reviewed the Defendants' Proposed ESI Order, and have participated in several meet and confers with the parties.

I. QUALIFICATIONS

3. I am a graduate of Stanford Law School, where I was a Note Editor of the Law Review. I was admitted to the bar in 1977 (I currently have retired status), and, after practicing law at Breed, Abbott & Morgan and Cravath, Swaine & Moore, I developed expertise in computer technology and software design, programming, and implementation, both generally and with respect to litigation support and e-discovery.

4. As an attorney at Cravath, I relied on *Aquarius*, the first large-scale implementation of computerized litigation support, which was implemented on the IBM antitrust cases.

5. As Director of Litigation Services at Legal Information Technology, Inc. ("LIT"), I was instrumental in introducing imaging, coding, and search technology for discovery to Am Law 200 law firms, and I pioneered the practice of integrating imaging with legacy search systems such as BRS.

6. As a systems architect, application designer and programmer, I created case management, litigation support and document repository systems and applications (including *WIDE*, and *LIT CaseWorks for Lotus Notes*, developed for LIT), SaaS (Software as a Service) knowledge management applications (including *LexisNexis Total Alerts* and *LexisNexis Clipper* developed for Ozmosys (https://www.ozmosys.com/)), and e-discovery and production operation systems for Doar (https://www.doar.com/). The software that I designed, programmed, and implemented to produce *LexisNexis Total Alerts*, *LexisNexis Clipper, LexisNexis Legal Industry Monitor, Thompson Elite Daily Docket, and Institutional Investors' Mutual Fund Daily, Hedge Fund Daily and Compliance Daily*, traversed thousands of web sites and extracted thousands of URLS from those sites daily.

7. At ILS, I direct e-discovery analytics and strategy, providing technical advice, expertise and drafting and other assistance to counsel.

8. I have advised, consulted, or acted as a declarant or affiant with respect to ESI in many cases, including:

a. *In re: Social Media Adolescent Addiction/Personal Injury Products Liability Litigation*, MDL No. 3047 (N.D. Cal.);

b. *In Re: StubHub Refund Litigation*, Case No. 4:20-md-02951-HSG (N.D. Cal.);

c. *In Re: Meta Pixel Healthcare Litigation*, Case No. 3:22-cv-3580-WHO (N.D. Cal.);

d. *In Re: Philips Recalled CPAP, Bi-Level Pap, and Mechanical Ventilator Products Litigation*, MDL No. 3014 (W.D. Pa.);

e. *Nichols v. Noom Inc.,* No. 20-CV-3677 (LGS) (KHP), 2021 WL 948646, (S.D.N.Y.);

f. *In re: Ethiopian Airlines Flight ET 302*, Lead Case: 1:19-cv-02170 (N.D. Ill.) (Boeing 737 Max Crashes);

g. *In Re: 3M Combat Arms Earplug Product Liability Litigation*, MDL No. 2885 (N.D. Fla.) ("3M Combat Arms MDL");

h. *In Re: Volkswagen "Clean Diesel" Marketing, Sales Practices and Products Liability Litigation*, MDL No. 2672 (N.D. Cal.) ("VW Clean Diesel MDL");

i. *In Re: Intel Corp, CPU Marketing, Sales Practices and Products Liability Litigation*, MDL No. 2828 (D. Or.);

j. *In Re: Chrysler-Dodge-Jeep EcoDiesel Marketing, Sales Practices and Products Liability Litigation*, MDL No. 2777 (N.D. Cal.);

k. *Lafferty v. Alex Jones* (Conn. Super. Ct.);

l. *Soto v. Bushmaster Firearms International* (Conn. Super. Ct.) (Sandy Hook parents);

m. *In Re: Takata Airbags Product Liability Litigation*, MDL No. 2599 (S.D. Fla.);

n. *In Re: Testosterone Replacement Therapy Products Liability Litigation*, MDL No. 2545 (N.D. Ill.);

o. *In Re JCCP 4771, Zoloft Birth Defect Cases*, (Cal. Super. Ct.);

p. *Da Silva Moore v. Publicis Groupe & MSL Group,* No. *11* Civ. 1279 (ALC)(AJP) (S.D.N.Y.) (seminal TAR case).

9. I have served as a speaker or panelist on many CLE webinars including *Balancing* the *Needs of Requesting and Producing Parties: Getting E-Discovery Right* (RAND Institute for Civil Justice Conference, October 3-4, 2023) ("RAND Conference"), *An Analysis of Today's Mass Tort Landscape Agenda* (HarrisMartin MDL Conference, March 27, 2019), *Current Mass Torts*

from E-Discovery Through Exit Strategies - Navigating "Game-Changing" Dynamics (HarrisMartin MDL Conference, November 26, 2018),*The Mass Tort Litigation Landscape - A Critical Analysis Agenda* (HarrisMartin MDL Conference, September 26, 2018), *The State of E-Discovery in 2018: Analysis & Review* (West LegalEdCenter, September 27, 2018), *Lessons Learned from Recent eDiscovery Disasters* (West LegalEdCenter, February 26, 2018), and *Top ESI Mistakes Made in Mass Tort Disputes* (West LegalEdCenter, September 14, 2017).

10. Along with Uber's expert Dr. Maura Grossman, I was an advisor on and assisted in setting up the RAND Conference.

11. I am a member of the drafting team for forthcoming commentary of the Sedona Conference Working Group 1 (https://thesedonaconference.org/wgs/wg1) on *Conducting eDiscovery of Modern Communication and Collaboration* and was on the panel presenting the *Drafting Team Report: Discovery of Modern Communications and Collaboration Platforms* at the Sedona Conference Working Group 1 Annual Meeting in October 2023.

II. OPINIONS

12. TAR 2.0, also known as CAL (continuous active learning), builds a predictive model based upon classifications of individual documents as responsive or non-responsive by human reviewers. The TAR 2.0 platform to be used here is Relativity Active Learning where the TAR process begins with the classification of an initial set of documents, which Relativity recommends be broadly selected to cover all relevant aspects of the matter, following which the TAR platform will continue to feed documents to the reviewers for classification as responsive or non-responsive. The classifications are then fed back into the predictive model which then re-scores the documents in the collection from 0 to 100 based on the predictive model's assessment of the responsiveness of each document.

13. By the conclusion of the TAR 2.0 process, every single document identified for production will have been classified as responsive by a human reviewer or be a family member of

such a document. The TAR 2.0 process continues until a stopping point, here defined as a drop-off in the percentage of documents classified as responsive by reviewers in two consecutive batches of documents to 10% or less, is reached. Once the stopping point has been reached, the TAR process is validated through classification of a set of sample documents.

14. If Plaintiffs do not participate in the key steps of the TAR process – creation of the initial set of training documents (the sample set), determination that the stopping criteria have been met and validation – then it would be more probable that the TAR program's predictive model will have been trained too narrowly, increasing the probability of downstream disputes over the adequacy of production, with expensive and time intensive motion practice and consequent schedule disruption and delays.

A. The TAR Protocol

1. *The Sample Set*

15. Relativity Active Learning, Defendants' chosen TAR platform, requires "at least five documents coded with the positive [responsive] choice and five coded with the negative choice" to begin training.[1]

16. Relativity recommends "having reviewers code a sample of documents prior to starting your Active Learning project to help start the [TAR] model's learning and begin the project with a model build and a ranking of all documents."[2]

[1] https://help relativity.com/Server2022/Content/Relativity/Active_Learning/Creating_an_Active_Learning_project.htm#:~:text=at%20least%20five%20documents%20coded%20with%20the%20positive%20choice%20and%20five%20coded%20with%20the%20negative%20choice. (Last accessed February 12, 2024)

[2] https://help relativity.com/Server2022/Content/Relativity/Active_Learning/Creating_an_Active_Learning_project.htm#:~:text=having%20reviewers%20code%20a%20sample%20of%20documents%20prior%20to%20starting%20your%20Active%20Learning%20project%20to%20help%20start%20the%20model%27s%20learning%20and%20begin%20the%20project%20with%20a%20model%20build%20and%20a%20ranking%20of%20all%20documents. (Last accessed February 12, 2024)

17. Relativity's recommendations for identifying documents to be pre-coded make it clear that the number of pre-coded documents should be much larger than only the minimum required 10 documents:

> "There are a few ways to identify documents to pre-code, and you will usually combine more than one strategy:
>
> - Draw a sample of documents with specific keywords, from key custodians, within a certain date range, etc. to help focus the sample on documents more likely to be important in the case. *Repeat for different aspects of responsiveness to best train the model*.
> - Cluster your documents, then use the choice field stratified sampling script to draw an equal number of documents from each cluster. *This ensures your sample will include a broad variety of documents*.
> - Use pre-existing documents determined to be relevant/not relevant."[3]

18. Defendants' proposed protocol is silent on the question of how the documents required to start will be selected and if, as Relativity recommends, Defendants will take measures to assure that this initial training set will include documents indicative of "*different aspects of responsiveness"* and include "*a broad variety of documents."*

19. The Sample Set proposed by Plaintiffs is a robust implementation of Relativity's recommendation, to "be created by drawing a simple random sample of 1,750 documents from the collection to which TAR will be applied." Exhibit 1, Section 8.3.ii.

20. In addition to its role in initiating the Relativity TAR training process, a Sample Set can also serve as a basis, through having both sides separately review and classify all non-privileged documents and then meet and confer to resolve any disagreements on documents where their assessments of relevancy differed, of helping assure that the parties' conceptions of relevance and scope are as congruent as possible at an early stage of discovery.[4] I think that one cannot

---

[3] https://help relativity.com/Server2022/Content/Relativity/Active_Learning/Creating_an_Active_Learning_project.htm#:~:text=There%20are%20a%20few%20ways,create%20your%20Active%20Learning%20project (emphasis added, cross reference link removed) (Last accessed February 12, 2024)

[4] An additional benefit of the sample is that it could serve as the basis for a richness assessment, which could be used

overestimate the importance of this process at this TAR stage in reducing the possibility of significant downstream schedule disruptions (and attendant expenses) when substantial remediation measures are required because producing parties were unable to view their responsiveness classifications from the requesting party's perspective.

21. This was the explicit reason for the sample sets in the *3M Combat Arms MDL* TAR Protocol, where, like in this MDL, Relativity Active Learning was the TAR platform: "The Parties' TAR approach will include the use of a Sample Set to help assure that the Parties have a sufficient level of agreement on what constitutes responsiveness and non-responsiveness." *In Re 3M Combat Arms Earplug Prods. Liab. Litig.,* No. 19-md-2885 (N.D. Fla. Jul. 1, 2019), Dkt. 472 at ¶ 3(a).

22. I played an active role in assisting Plaintiffs' counsel in drafting the 3M TAR Protocol and, as Relativity Active Learning, used in 3M is also to be used here, the 3M language on sample sets and TAR training was directly incorporated into those sections of Plaintiffs' proposed protocol in this matter.

23. A similar function, helping to align the parties' conceptions of relevance and scope at an early stage of the litigation, was also performed by the use of a Control Set in the Volkswagen Clean Diesel MDL before Judge Breyer which I also assisted Plaintiffs' counsel in drafting. The Control Set in Volkswagen was a simple random sample of 4,000 documents and as in 3M and proposed by Plaintiffs here, all of the non-privileged documents in the sample were to be separately reviewed and classified as responsive or not by both the producing and requesting parties, with any disagreements to be resolved via meet and confer and if necessary by the court.

24. Serving as a Special Master in the seminal case of *Rio Tinto v. Vale*[5], Uber's expert Dr. Grossman stipulated order directed that all non-privileged documents in a control set and in "Audit Samples" (essentially validation samples) be provided to the requesting party with any disagreements on the classification of specific documents to be resolved by meet and confer and in

as an additional yardstick against which the total number of responsive documents can be compared.

[5] *Rio Tinto plc v. Vale S.A.,* No. 14-cv-3042, 2015 WL 13956130, at *2 (S.D.N.Y. Sept. 8, 2015)

the case of disagreements, by the Special Master and, if appropriate, the Magistrate Judge. *Rio Tinto* Stipulation and Order Re: Revised Validation and Audit Protocols for the User of Predictive Coding in Discovery Paragraphs 2(c)-(d), 5(a)-(e).

2. *The TAR Training Process*

25. Defendants' description of their proposed TAR training process is quite skeletal, contained in a single paragraph:

> "As part of document review, the Uber Defendants intend to use TAR methodology known as TAR 2.0, which utilizes continuous active learning to classify and prioritize documents for attorneys to review. Specifically, the Uber Defendants intend to use Relativity Active Learning ("RAL") on a Relativity Server 12.1.537.3 platform provided by their vendor Lighthouse. Commonly, a TAR 2.0 methodology begins with ingesting document population into the TAR 2.0 software where the algorithm learns to distinguish relevant from non-relevant documents through attorney review of documents. The TAR 2.0 algorithm prioritizes the documents in the review queue in a more efficient manner. Attorney reviewers then review documents the TAR 2.0 model has prioritized as most likely to be responsive. As the review continues and reviewers code documents, the TAR 2.0 model continues to learn and prioritize likely responsive documents until a stopping point is reached and a validation is conducted." Exhibit 2 Section 8.a)1.

26. This barebones description omits many critical details:

a. Although, given the discussion of batches in the next section (8.2) of Defendants' proposal on stopping criteria, it appears that Defendants intend the review to be conducted in batches, batches and the size and composition of batches are unmentioned in Defendants' description of their TAR training process.

b. Defendants state that "Attorney reviewers then review documents the TAR 2.0 model has prioritized as most likely to be responsive." This, however,

> does not provide information about the construction or size of the batches (such as whether the highest-ranking uncoded documents remaining in the project in order from highest to lowest scores are used exclusively). Or, does the reference to "prioritized" review indicate that at some point in the process, the review will utilize Relativity's Prioritized Review queue which "serves up a mixture of documents: 10% of documents are randomly selected; 20% of documents are chosen for scores "in the middle" of the review (in the 40 to 60 range) for index health; and the final 70% are the highest-ranking uncoded documents remaining in the project. This means for every 200 documents, [only] around 140 (70%) are chosen for being highly ranked"[6]

27. I note this because of Defendants' statement in its description of its proposed stopping criteria procedure states that "Defendants do not intend for the *relevant batches* to include index health documents" Exhibit 2, Section 8.2.i (emphasis added). This implies that earlier batches may, like Relativity's Prioritized Review queue, incorporate index health documents.

28. This is of concern because of conclusions reached by Dr, Grossman in her recent co-authored paper *Comparison of Tools and Methods for Technology-Assisted Review*[7] ("Comparison"). This study, *inter alia,* compared results reached by Grossman and Cormack's CAL® system against an unnamed "leading eDiscovery provider's TAR tool" which, although unnamed in the study, is evidently the Relativity Active Learning tool that Defendants intend to use here.

---

[6] https://help relativity.com/Server2022/Content/Relativity/Active_Learning/Prioritized_review.htm#:~:text=The%20Prioritized%20Review%20queue%20serves%20up,are%20chosen%20for%20being%20highly%20ranked. (Last accessed February 12, 2024)

[7] Halloran, McManus, Harbison, Grossman, & Cormack, Comparison of Tools and Methods for Technology-Assisted Review, https://ediscoverytoday.com/wp-content/uploads/2024/01/Comparison-of-Tools-and-Methods-for-Technology-Assisted-Review.pdf (Last Accessed February 12, 2024)

29. In the Comparison study, Dr. Grossman and her co-authors found that use of what appears to have been Relativity's Prioritized Review queue "increases review effort considerably," "may contribute to why the provider's tool [Relativity Active Learning] underperformed regarding the recall it achieved," and "can give rise to substantial duplication of review effort and not only protract the length of time it takes to complete a review but incur additional expense on the user."[8]

30. If Defendants do use Relativity's Prioritized Review queue, even if not in the final "*relevant batches*" to be used for determining if their stopping criteria has been reached, then Defendants may incur avoidable and unnecessary expenses which Defendants could adduce in opposing the remedial measures that may be required if TAR is prematurely stopped because of the inadequacy of their stopping criteria, or if their TAR process passes validation only because Plaintiffs, who may have a far different conception of relevance than Defendants, were not included in the validation process.

3. *Stopping Criteria*

31. Stopping criteria will determine, subject to subsequent validation, when TAR 2.0 training, which will have incorporated the manual responsive/non-responsive classification of every document to be produced (or logged as privileged), can stop. Defendants stopping criteria is based on the occurrence of a drop-off in the proportion of responsive documents found in an assessment set to 10% or fewer. Plaintiffs do not object to this methodology as long as the number of documents in the assessment set is sufficient to minimize the possibility of a false positive, i.e., a determination that the percentage of responsive documents in the review queue has fallen persistently below the 10% cutoff, when in fact the drop in the proportion of responsive documents below the 10% threshold is only temporary, and the longer-term proportion of responsive documents will subsequently return above this cutoff.

32. Defendants stopping criteria, in relevant part, is:

[8] *Id.* at 7.

> "Once two reasonably sized review batches are found to contain 10% or fewer documents marked responsive, Defendants will pause the review and turn to validation. Defendants may extend the review past this point if they believe sufficient thoroughness has not been achieved. Defendants do not intend for the relevant batches to include index health documents." Exhibit 2 Section 8.a.2.i.

33. Relativity permits the size of batches to be determined by the party using it (my understanding is that batch sizes of 200 or 250 documents are not uncommon), but Defendants have thus far refused to specify what the number of documents in a reasonably sized batch would be and consequently what the total size of the assessment set, which would consist of two of these batches, would it be. Would it be 100 documents per batch for a total of 200 documents? Or 200 documents per batch for a total assessment set of 400? Or some other number of documents per batch? Defendants have refused to provide this basic information.

34. Plaintiffs have accepted Defendants' use of this type of stopping criteria which is a measure of marginal precision[9], but have set 1,000 as the number of documents in which the percentage of responsive documents will be measured:

> "The Uber Defendants will continue to review the prioritized review queue until two or more consecutive batches containing a total of at least 1,000 documents prioritized as likely responsive are each found to contain 10% or fewer documents marked responsive. Defendants will disclose that the stopping criteria has been met, and, for each batch, the number of documents and responsive documents in it. Defendants may extend the review past this point if they believe sufficient thoroughness has not been achieved."

35. Plaintiffs express no views at this time on a preferred size of the stopping criteria batches or how many batches there are so long as there are at least two batches and the number of batches multiplied by the number documents in each batch totals at least 1,000.

[9] "Precision," as used here is a statistics metric which in this context means the percentage of responsive documents in a given collection or sample.

36. Differences in the size of the assessment set can result in drastic differences in the probability that evaluation of the percentage of responsive documents will result in a false positive, i.e., the training process will be stopped prematurely. In order to illustrate this, I directed analysis on assessment sets consisting of 200 documents (equivalent to two batches of 100 documents each), 400 documents (equivalent to two batches of 200 documents each) and 1,000 (equivalent to five batches of 200 documents or four batches of 250 documents, etc.).

37. Among the conclusions reached in the analysis are that:

a. If you stop too early and review repeat testing and evaluate the effect of the "sequential testing" problem, the probability of a false positive for an assessment set of 200 documents is 72% while the probability of stopping too early with an assessment set of 400 would be a non-negligible 5.2% and which would drop to only 0.01% (1 in 10,000)[10] for an assessment set of 1,000.

b. A further complication would be the effects of the cluster hypothesis, a key tenet of information retrieval that holds that states that responsive documents tend to be more similar to each other than non-responsive documents. If responsive documents tend to come in sequences, it follows that (for a given responsive rate), non-responsive documents will also tend to come in sequences. Using a simplified model[11], the probability of

[10] Per the analysis, sequential testing can lead to serious estimate biases if not correctly accounted for. Crucially, here the review and testing continues if a test fails (the threshold is not met), but stops as soon as a single test passes. Variance overstating the responsiveness rate is not penalized, but random variation understating it is rewarded. As a result, the probability of stopping too early grows with the number of tests performed.
Returning to the simple random model used in the analysis, if each individual test has an independent probability of p of stopping too early, then the probability of stopping too early after k tests is $1 - (1 - p)k$, an instance of the geometric distribution.
Consider the case of testing a sequence of 10,000 documents drawn from a span with a 15% responsiveness rate. If the assessment set size is 200, then we will perform 50 tests (or until one passes). Given that each test has a 2.5% chance of passing, the probability of a passed test and a premature stop is 72%. We are far more likely than not to stop early. In contrast, if the assessment set size is 1,000, then we perform 10 tests. Given that each test has a 0.0002% chance of passing, the probability of one passed test in the sequence is 0.01% (one in ten thousand). Therefore, even with the sequential testing bias, the probability of stopping too early would be negligible.

[11] Per the analysis, and put in practical terms, what we will tend to observe in CAL review queues, particularly as the responsiveness rate drops, is sequences of similar responsive documents, followed by (perhaps extended) sequences in

stopping too early would be 97.2% for an assessment set of 200 documents, 47.5% for an assessment set of 400 documents and only .65% for an assessment set of 1,000 documents.

38. I also note that in a prior litigation, Dr. Grossman and I assisted in drafting an unadopted ESI protocol including, as in this matter, the use of Relativity's TAR 2.0 functionality. In drafts approved by Dr. Grossman, the stopping criteria was based on the number of responsive documents found in the 1,000 most recently reviewed documents in the five most recently reviewed groups of 200 documents.

4. *TAR Validation*

39. Plaintiffs proposed a robust and reasonably transparent validation process, including key metrics on the process, review by Plaintiffs of all non-privileged validation sample documents, production of all non-privileged responsive documents and a meet and confer to resolve any disputes over the responsive/non-responsive classification of individual documents and subsequent court resolution of any remaining disputes. Exhibit 1, Section 8.a.6,and 8.b.

40. Defendants proposed protocol does not allow Plaintiffs any role in the validation process[12] – no separate review of any non-privileged documents, no meet and confer to resolve disagreements over the responsiveness of individual documents, no recourse to the court to resolve

---

which no responsive documents are found—only to be followed by other sequences of responsive documents. This tendency increases the likelihood of stopping too early, if the assessment set size is small and a given assessment set happens to fall in a predominantly non-responsive sequence.
A full general random model of a CAL review under these conditions of non-independence is beyond the scope of our analysis. The direction of effect, however, can be illustrated by a simplified model. Assume that responsive and non-responsive documents come strictly in sequences of length 2, and that the appearance of responsive (and non-responsive) sequences are independent of each other. Then a responsiveness rate test with an assessment set size of 200 is equivalent to a binomial test with a 200 ÷ 2 = 100. Given a true rate of 15%, the probability of incorrectly passing an individual test with a threshold of 10% is 9.9%. Furthermore, the probability of passing at least one of the 50 such tests in a sequence of 10,000 documents is 99.5%—in other words, it is essentially certain that we will stop too early. In contrast, if the assessment set length is 1,000, then the probability of passing an individual test is only 0.065%, and the probability of passing at least one of the 10 such tests in a sequence of 10,000 documents is 0.65%. In other words, increasing the assessment test size from 200 to 1,000 reduces the risk of early termination, given the clustering of responsive and non-responsive documents, from almost certain to negligible.

[12] Defendants' proposed protocol does have a provision that "the parties may meet and confer to discuss reasonable questions and issues relating to the TAR process" *after the TAR process has concluded.*

any disagreements not resolved through the meet and confer, and no attendant production of the non-privileged responsive documents.

41. As in Plaintiffs' proposal, the TAR protocols in the *3M Combat Arms MDL* and the *Volkswagen Clean Diesel MDL* and Dr. Grossman's *Rio Tinto* protocol all allowed Plaintiffs to review all non-privileged documents in the validation sample, to dispute the responsive classification of individual documents in a meet and confer, to have remaining disagreements resolved by the court and to receive the non-privileged responsive documents in the process.

42. Dr. Grossman's orders as Special Master in *In re Broiler Chicken Antitrust Litig.*, No. 1:16-cv-0637 (E.D. Ill. Jan. 3, 2018) and *BCBSM, Inc. et al.* v. *Walgreen Co. and Walgreens Boots Alliance, Inc.*, No. 1:20-cv-01853, 2023 WL 6852533 (N.D. Ill. July 21, 2023) (referenced in her JCCP declaration) allowed Plaintiffs to review all non-privileged documents classified as responsive by the producing party in the validation sample, to dispute the responsive classification of individual documents in a meet and confer, to have remaining disagreements resolved by the court and receive the non-privileged responsive documents in the process.

43. *Broiler Chicken* described the meet and confer process like this:

> "Once the requesting Party has received and has had an opportunity to review the items described in Paragraph III(H) and Appendix A[13], the Parties shall meet and confer to determine whether or not the Parties agree that the recall estimate, and the quantity and nature of the responsive documents identified through the sampling process, indicate that the review is substantially complete. If the recall estimate and the samples indicate that Subcollections C(2) and/or C(3) still contain a substantial number of non-marginal, non-duplicative responsive documents as compared to

[13] The described items include (1) a table listing, *inter alia,* the Bates number of the document (for documents produced), or a control/identification number (for non-produced documents); the Subsample from which the document came (i.e., D(1), D(2), or, if TAR was used, D(3)); the reviewing Subject Matter Expert(SME)'s responsiveness coding for the document (i.e., responsive or non-responsive); the SME's privilege coding for the document (i.e., privileged or not privileged) and for putative class plaintiffs, the named class representative associated with the document (2) a copy of each responsive, non-privileged document in the Validation Sample that was not previously produced or identified for production to the requesting Party and (3) the statistics and recall estimate detailed in Appendix A of the Order.

> Subcollection C(1), the review and quality assurance process shall continue, and the validation process shall be repeated, as warranted."
>
> *In Re: Broiler Chicken,* Section III.I.

44. Validation processes and procedures that include meaningful participation by the requesting parties, like those set out in Plaintiffs protocol and in the just-cited cases are essential to ensure that an entirely unilateral validation process does not cut off discovery prematurely, setting the stage for disruptive, expensive, and schedule killing disputes and consequent remedial measures downstream.

5. *End-to-End Validation*

45. Recall in this context is the percentage of responsive documents correctly identified as being responsive in discovery. It is a fundamental metric for judging the adequacy and reasonableness of discovery.

46. While all Parties recognize the importance of recall, they disagree with how it is to be measured with respect to some aspects of Defendants' proposed validation process as discussed in paragraphs 31-38 of this declaration, and whether recall should be calculated just for TAR or also for the pre-TAR application of search terms and for Defendants' end-to-end discovery process including manual review.

47. Defendants have rejected calculating recall on the results of their entire end-to-process and would calculate recall only for its proposed TAR process, even though they apparently intend to impose at least two additional steps in their end-to-process, viz, filtering their collection with search terms before TAR is applied and manually reviewing the documents identified as responsive by TAR after TAR.

48. Suppose only 20% of the universe of a party's responsive documents were identified as responsive during discovery. That would plainly be inadequate under any circumstances.

49. Further suppose that this 20% recall was the result of a multi-step process that includes search term filtering followed by TAR followed by manual review, and that the recall for just the TAR portion was 80%. That 80% TAR recall would be from a population that had already had a substantial portion, perhaps 70% or higher, of its responsive documents removed by the prior application of search terms.

50. Accordingly, when multiple processes are applied in winnowing down a collection, the relevant recall rate is the product of the recall rates of each process, not the recall rate for just one process.

51. As Defendants' expert Dr. Grossman explained in her co-authored article *The eDiscovery Medicine Show*:

> "To make matters worse, eDiscovery medicine shows frequently promote the sequential use of two or more information retrieval methods, including Boolean search, TAR, and manual review. *The net effect of this concoction is to achieve considerably lower recall than any of its constituent parts: when multiple information retrieval methods are used in sequence, overall recall is the product of the recall for each constituent method.* If keyword culling were to achieve 70 percent recall, the TAR tool were to achieve 80 percent recall, and manual review were to achieve 75 percent recall, the recall of a review effort combining them in sequence would be 70 percent × 80 percent × 75 percent = 42 percent. It is possible to quibble with the numbers presented here, but not with the fact that each constituent part is imperfect, and that overall or end-to-end recall is considerably less than the weakest link in the chain.
>
> "When applied sequentially, information retrieval methods—whether Boolean search, TAR, or manual review—will always yield inferior recall. Yet the medicine shows would have us believe that we need to consider only the TAR-tool component in our recall calculations, ignoring relevant documents excluded by keyword culling and/or by post-TAR manual review. This is, at best, an extreme case of moving the goalposts, but more

likely, a form of legerdemain.""

Maura Grossman & Gordon Cormack, *The eDiscovery Medicine Show,* 18 The Ohio State Technology Law Journal 1, 8 (1921) (footnotes omitted)[14]

52. Dr. Grossman's declaration in *In re Diisocyanates Antitrust Litig.,* No. 18-mc-01001 (W.D. Pa. Mar. 25, 2021), Dkt. 459 at ¶ 35 *et seq.* also makes this point:

35. I understand that the parties have agreed to use search-term pre-culling, which will result in reduced recall even before the TAR process begins. An accurate and proper measure of recall evaluates the totality of the review effort—including the impact of the keyword culling, TAR, and human error [in post-TAR manual review]—and therefore, is based on the entire document collection, not just the TAR review set.

36. *Validation (and recall calculation) should apply to the end-to-end search and review process, starting with the original collection and ending with the production set, regardless of which population TAR is applied to. That is, validation must account for all responsive documents potentially excluded by the search and review process—as well as non-responsive documents included in the production set. Calculating recall only on a smaller body of documents, or a single phase of the review process, necessarily diminishes the validity of the recall measure and misrepresents the actual quality of the production.*

37. To make this multiplicative impact more concrete, imagine a total collection of one million documents of which 100,000 are responsive. Further imagine that 60% of the documents are eliminated through search-term pre-culling, such that 400,000 documents move on to TAR. Let us further assume that 70,000 of the 400,000 documents are responsive. It follows that 30,000 responsive documents would be left behind among the

[14] Available at https://moritzlaw.osu.edu/sites/default/files/2022-01/THE%20EDISCOVERY%20MEDICINE%20SHOW.pdf.

600,000 documents excluded by the search terms. That is, the recall of the search terms would be 70%. (In my experience, search-term recall would typically be lower than 70%, since search terms often tend to miss more responsive documents than 30%.) If Defendants were to achieve their "target recall" of 70% on the TAR review population of 400,000 documents, they would stop the review when they found 49,000 responsive documents. But the recall of the search and review process would not be 70% as suggested; it would actually be 49% (70% x 70%), because of the prior application of the search terms. And even that would be an exaggeration because the human reviewers would miscode a certain number of responsive documents as non-responsive. Let's assume the reviewers correctly code 70% of the documents, which is consistent with empirical research and my own experience. Then, the actual recall of the production set would be even lower (70% x 70% x 70% = **34.3%**). That is, only 34,000 of the 100,000 responsive documents would be produced. *Even using a broad set of search terms, which might return a greater percentage of responsive documents—let's say 85%—the recall would still drop substantially when calculated for the whole end-to-end process* (85% x 70% x 70% = **41.65%**). Paragraphs 35-37 (citation omitted; some emphasis added).

6. *Key Word Search*

53. The deficiencies of search terms in litigation were exposed in the 1985 publication of the landmark paper by Blair and Maron[15] that found that only 20% of relevant documents included search terms selected by experienced counsel (who thought that the search terms would have retrieved 75% of relevant documents).

[15] David C. Blair & M. E. Maron, An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System, 28 COMMC'NS ACM 289, 290 (1985).

54. Later studies at the 2007 and 2008 TREC Legal Track[16] confirmed this dismal state of affairs, finding estimated recall rates for search terms of 22%, i.e., 78% of relevant documents missed, and 24%[17], missing 76% of relevant documents. Perhaps worse, with respect to highly relevant documents, the result of the application of a "consensus set" of Boolean search terms in the 2008 Legal Track implied "that, on average per topic, *58% of the 'highly relevant' documents were not found* … indicating that it is not just tangentially relevant documents that are being missed by the negotiated Boolean approach.[18]"

55. Because of the dismal recall of search terms and its consequent drastic effects on end-to-end recall, Plaintiffs' proposed protocol bars the use of search terms prior to TAR and, unless required as a post-TAR remedial process, after TAR. Exhibit 1, Section 9.a.2.

56. Whether search terms are used instead of TAR, or using search term before TAR is permitted, the recall of the search terms should be calculated. In the former case, to ensure that the search terms capture a reasonable proportion of responsive documents and in the latter case to enable accurate calculation of end-to-end recall. Accordingly, Plaintiffs' proposal includes a provision for calculating recall and validating the search process using a methodology and metrics comparable to those used to calculate recall and validate TAR. Exhibit 1, Section 9.c.

B. *Collection of Hyperlinked Documents*

57. Hyperlinked documents, also called modern attachments, serve the same purpose and role as traditional attachments, especially in environments such as Google Workspace, where Google Mail is the email platform and the common business file types are native Google file

[16] The Text Retrieval Conference (TREC), co-sponsored by the NIST Information Technology Laboratory's (ITL) Retrieval Group was established in 1992. Celebrating 25 Years of TREC, available at https://trec nist.gov/celebration/25thcelebration.html. The Legal Track at the Text REtrieval Conference (TREC) was established "to assess the ability of information retrieval techniques to meet the needs of the legal profession for tools and methods capable of helping with the retrieval of electronic business records, principally for use as evidence in civil litigation." TREC Legal Track, available at https://trec-legal.umiacs.umd.edu/

[17] Stephen Tomlinson, et al., Overview of the 2007 TREC Legal Track (April 30,2008), (22% recall), available at https://trec nist.gov/pubs/trec16/papers/LEGAL.OVERVIEW16.pdf; Douglas W. Oard, et al., Overview of the TREC 2008 Legal Track (November 1, 2008), (finding 24% recall), available at https://apps.dtic.mil/sti/pdfs/ADA512711.pdf.

[18] *Oard, supra,* at p. 39 (emphasis added).

types, with Google docs standing in for Microsoft Word, Google sheets standing in for Microsoft Excel, and Google Slides standing in for Microsoft Power Point.

58. But a critical difference between Google native file types and the corresponding Microsoft file types is that, unlike their Microsoft analogues, Google native file types ("Google documents") cannot exist outside the Google environment, i.e., a native Google file type cannot be saved as an individual file and attached to an email like a Microsoft document can be.

59. Companies like Uber choose to use Google for email and documents must for virtually all practical purposes[19] use hyperlinks to include Google documents as attachments and would be well aware of that limitation when they opted to use Google as their platform.

60. In discovery of email and documents with traditional attachments, the parent documents are linked with attachments in parent-child relationships and family groups; when a responsive document is produced, its entirely family will as a rule also be produced[20], regardless of whether the family member identified as responsive was a parent or an attachment.

61. This is especially critical in cases where, unless the sender of an email was the author or otherwise inherently linked to an attachment, the relationship between an attached document and a parent email and its sender and recipients, important substantive evidence in its own right, may usually have to be established through the evidence of that parent-child relationship.

62. Moreover, traditionally attached documents will always be the version that existed at the time that the parent email was sent.

63. Using Google Workspace blows all this up. Because Google documents cannot exist outside of the Google environment, they must be converted into file types that can exist outside of the Google environment, i.e., their Microsoft analogues.

[19] Unless corporate policy forbids it, a user could export a document from Google Drive as its Microsoft equivalent and attach that to an email.

[20] Single page image placeholders are usually produced if a family member has been withheld from production.

64. For discovery purposes, in order to preserve the family and evidentiary relationships and evidentiary value of Google Workspace emails and documents, two critical requirements must be met.

65. First, the parent-child/family relationships that existed in the Google environment must be preserved, i.e., the exported Microsoft version of a hyperlinked attachment must be linked to the exported version of the email that contained that hyperlink in Google.

66. Second, that exported version must be the version that was current at the time that the email was sent.

67. Defendants' proposed process is to use Google Vault to collect documents and emails and export them for discovery and their vendor Lighthouse's proprietary "Google URL Parser" to link them.

68. The use of Google Vault is problematic because, even though Google Vault does generally preserve the prior versions of a document[21], as of now, the version of a document exported by Vault will always be the most recent version as of the time of export even though third-party programs such as Metaspike's *Forensic Email Collector* can do exactly that, i.e., export the version of a hyperlinked document that existed as of the time that an email was sent[22].

69. Second, Defendants' proposed provision on hyperlinked attachments (Exhibit 2, Section 17), with its reference to "technological limitations" glosses over the fact that Defendants chose this technology and their eDiscovery vendor. Additionally, Uber is doing their collection in-house.

70. *Forensic Email Collector* has been used for years by demanding organizations. Prominent FEC users, including users in government and law enforcement, and in prominent firms in the Am Law 100, the accounting Big 3, and litigation support and forensic services, include Paul Weiss, PriceWaterhouseCoopers, the Office of the California Attorney General, FTI

[21] Early versions of very large files—except those subject to a hold—might be merged by Google Vault to save storage space.

[22] Up until December 8, 2023, Google Vault did not even have the capability to collect hyperlinked document attachments and associate them with their parent emails,

Consulting, Deloitte, Grant Thornton, Clifford Chance, Stroz Friedberg, Winston & Strawn, the Federal Trade Commission, the Norwegian Tax Administration, and the Australian Federal Police. (Battle-Tested Software, https://www.metaspike.com/forensic-email-collector/#customers, accessed on March 7, 2021).

71. Defendants have objected to the use of *Forensic Email Collector* on some general grounds including that they are concerned about scale and security about using a third-party vendor and did not believe it can access emails that are stored in Google Vault (Uber's document retention system) and thus will be incapable of retrieving emails/linked documents that are only accessible in Vault (which would be the case for older emails).

72. In response, I note the following:

a. *Forensic Email Collector*'s inability to access emails stored in Google Vault does not affect its ability to collect and process emails and documents that are *not* in Google Vault and would not have affected its ability to do the same earlier when documents that now exist only in Google Vault still existed outside it in Google Drive.

b. If Uber has not already tested or used *Forensic Email Collector*, any security concerns that Uber has could be resolved in the same way that Uber usually tests any other third-party software that Uber has introduced to run on its system. Moreover, Uber has decided to use a platforms which are run by third party vendors and may be subject to data leaks and security breaches, versus maintaining their information in-house on applications run internally on in-house physical servers.

c. Scalability issues can be caused by many different factors, including misconfiguration or sub-optimal application of a program, network configuration and congestion and throttling imposed by Google. Defendants' counsel have not provided any details – how many email accounts, how many copies of Forensic Email Collector deployed, network

and internet configuration, etc. – about the specific deployments of Forensic Email Collector with which they had issues. I note that ILS has reached out to Metaspike and were informed that multiple copies of Forensic Email Collector can be deployed and that it was aware of instances where Forensic Email Collector was faster than Google Vault.

73. What of the documents that now, because of Uber's policies on document retention and deletion and Google Vault deployment, exist only in Google Vault? Although Google Vault exports only the most recent version a document, the Google Vault environment does provide functionality to enable users to preview and export earlier versions[23]. Google Vault also has an API that should be explored. To the extent that Google Vault requires manual steps to recover a document as it existed at a particular date, macro recorders may enable automation and should also be considered.

C. *Deduplication Metadata*

74. Plaintiffs' provision on deduplication included this language:

"The names of all custodians and non-custodial sources who were in possession of a document prior to deduplication will be populated in the "ALL CUSTODIAN" metadata field, separated by semi-colons, in addition to a separate field of data identifying the custodian whose Document is produced; such de-duplicated Documents shall be deemed produced from the custodial files of each such identified custodian for all purposes in this litigation, including for use at deposition and trial. The original file paths of a Document prior to deduplication will be populated in the "ALL FILE PATHS" metadata field, separated by semicolons, in the order corresponding to the order of names in ALL CUSTODIANS. Hard-Copy Documents shall not be eliminated as duplicates of ESI."

[23] "You can search only the most recent version of a file. However, you can enter a version date as part of your search. When you enter a version date, Vault searches the current version of the file, but you preview and export the contents of the last version saved before 12:00 AM on the specified date. You set the time zone in the query. Google Drive retains versions of files created with Google Docs, Sheets, Slides, Sites, and Drawings." https://support.google.com/vault/answer/7654308#zippy=%2Csearch-within-the-revision-history-of-a-file:~:text=You%20can%20search%20only,Slides%2C%20Sites%2C%20and%20Drawings.

75. Defendants' version strikes the phrase "in the order corresponding to the order of names in ALL CUSTODIANS."

76. File paths, the location where custodians stored their copies of a file, are substantive evidence. For example, a custodian could have stored their copy of a file in a directory named "Rogue Drivers." Under Plaintiffs' language, Plaintiffs would be able to match an individual file path such as "Rogue Drivers" with the specific custodian who stored their copy of the document in that location, by matching the position of the file path in ALL FILE PATHS with the custodian in that same position in the ALL CUSTODIANS field. Defendants' language would strip Plaintiffs of that capability.

77. Defendants' rationale for stripping Plaintiffs of that capability is a claim that this is not part of their vendor Lighthouse's normal processing.

78. The phrase in question is common in ESI protocols and this is the first time in working on 100's of ESI protocols that I have seen any objection to it. Lighthouse should and could provide this capability. Defendants have not articulated the burden in producing this but merely stated in it not part of the standard processing format.

I declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge.

DATED: February 12, 2024

Douglas Forrest
Senior VP, eDiscovery
Analytics & Strategy
International Litigation
Services, Inc.
dforrest@ilsteam.com